

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/298350110>

Towards the Discovery of Scientific Revolutions in Scientometric Data

Article in *Scientometrics* · January 2017

DOI: 10.1007/s11192-016-2108-x

CITATION

1

READS

191

1 author:



Rogier De Langhe

Ghent University

30 PUBLICATIONS 212 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



The Economics of Economics [View project](#)

Towards the Discovery of Scientific Revolutions in Scientometric Data

March 31, 2016

Rogier De Langhe, Centre for Logic and Philosophy of Science, Ghent University, Belgium

Abstract

Paradigms and revolutions are popular concepts in science studies and beyond, yet their meaning is notoriously vague and their existence is widely disputed. Drawing on recent developments in agent-based modeling and scientometric data this paper offers a precise conceptualization of paradigms and their dynamics, as well as a number of hypotheses that could in principle be used to test for the existence of scientific revolutions in scientometric data.

Introduction

In his book *The Structure of Scientific Revolutions* Thomas Kuhn (1970), a Harvard condensed matter physicist (PhD 1949), described science as what could today be called a “complex adaptive system”. Kuhn described science as a process characterized by pre-paradigmatic, normal and revolutionary patterns emerging from the interactions of its component scientists.(De Langhe 2013) Today it is commonplace to study complex adaptive systems using combinations of agent-based simulation and very large datasets.(Miller & Page 2007) Yet in Kuhn’s time the tools, theories and datasets to articulate this view were still in their infancy. For lack of statistical data to study systemic patterns in science, scholars inspired by Kuhn (including the early Kuhn himself) took recourse to the historical record and looked for patterns in historical case-studies instead. This might explain why Kuhn’s popular image of science has mostly been critically received by professional philosophers. Secondary literature on Kuhn identifies key weaknesses in every major aspect of Kuhn’s view: no clear mechanism for rationality and progress is laid out (Sharrock & Read 2003), the core concepts allow for too many alternative interpretations (Masterman 1970) and evidential support is indirect: particular historical case studies are supposed to support claims about general patterns in science. Kuhn himself later even called historical case-studies “misleading” (Kuhn 2000, 111) because their study only deepens the problems they suggest rather than solving them. In a paper titled ‘The Trouble with the Historical Philosophy of Science’ he writes that “many of the most central conclusions we drew from the historical record can be derived instead from first principles”(Kuhn 2000, 112) and these principles “are necessary characteristics of any developmental or evolutionary process”(Kuhn 2000, 119). Kuhn even had plans to publish a

book on such an evolutionary theory of scientific discovery. Unfortunately Kuhn passed away and the book was never finished.¹ Although all but non-existent in his time, already in the second edition of *Structure* in 1970 Kuhn had asserted the relevance scientometric datasets could one day have for the verification of his claims.²

“for this purpose [the purpose of detecting paradigms] one must have recourse to attendance at special conferences, to the distribution of draft manuscripts or galley proofs prior to publication, and above all to formal and informal communication networks including those discovered in correspondence and in the linkages among citations. *I take it that the job can and will be done*” (Kuhn 1970, 177-8, my italics)

Although there is a vast secondary literature on Kuhn, most of it was written in the 1970s and 1980s. More recent contributions have interpreted Kuhn using concepts from philosophy (Hoyningen-Huene 1993), the cognitive sciences (Andersen, Barker, & Chen 2006), evolutionary biology (Wray 2011) or political sciences (Fuller 2001). The once massive interest³ in *Structure* began to stall in the 1990’s, paradoxically right at the time when scientometric datasets started to become widely available thanks to the digitization of scientific research and methods for the study of complex adaptive systems started to mature, for example at the Santa Fe Institute (Arthur 1994, Holland 1998). Save some exceptions, development of agent-based modeling and the widespread availability of scientometric datasets has only very recently sparked renewed interest in the work of Kuhn. Roughly this work is either theoretical (Chen 2012, Chen et al. 2009, Sterman 1985, Sterman & Wittenberg 1999) with a call on bibliometricians to operationalize their concepts^[4], or empirical (Moravcsik and Murugesan (1979), Radicchi, Fortunato, and Castellano (2008), Mazloumian, Eom, Helbing, Lozano, and Fortunato (2011), L. Bettencourt, Kaiser, and Kaur (2009), L. Bettencourt and Kaur (2011), M. Bettencourt and Kaiser (2015)) calling on theoreticians to provide a theoretical framework.^[5] ^[6] This paper takes an opposite approach. Instead of doing detailed theoretical work and calling on empiricists to complement it (or vice versa), I present in this paper a simple yet general approach with the potential to explain a number of macroscopic empirical patterns such as why scientific activity clusters in fields, why these fields are larger in some areas and smaller or even barely existing in others and why the structure of scientific fields changes through time. By keeping the model simple but general, it is conceived rather as a standard

¹Kuhn had appointed James Conant and the late John Haugeland to finish the manuscript. (Nickles 2003) Even twenty years after Kuhn’s death the book is still “forthcoming” according to James Conant and Chicago University Press.

²On page 178 Kuhn (1970) references an early book by his friend Eugene Garfield (1964) who would come to be known as one of the founders of scientometrics.

³The book sold more than 1 million copies making it one of the most widely distributed philosophical books in the world.

⁴*Bibliometric techniques could be used to determine how long a research problem (“puzzle”) has gone unsolved and gauge the number of researchers working on it, to yield a measure of the difficulty of puzzles.* (Sterman & Wittenberg 1999)

⁵*Finally, from a more theoretical point of view, an interesting goal for future work is to understand the origin of the universality found and how its precise functional form comes about.* (Radicchi et al. 2008)

⁶*The existence of a general theory and detailed model that describes the formation of scientific fields across disciplines, time, and population size would provide a new comprehensive, quantitative, and predictive framework with which to understand the social and conceptual dynamics that drive the self-organized creation of scientific communities. Such a framework would be of significant interest to scientists and would hold great promise for guiding science policy.* (M. Bettencourt & Kaiser 2015)

for further theoretical and empirical refinement within a framework that already connects the theoretical and empirical. This approach is a result of my broader, mixed ambition to find evidence for the existence of scientific revolutions in scientometric data. I believe this would constitute a major theoretical result achieved by empirical means.

The paper will consist of two parts. In the first two sections I build an agent-based model to show how the interactions of individual scientists could possibly result in Kuhnian patterns of science. In the second part I turn to the question how the results of this agent-based model could in principle be operationalized to hunt for scientific revolutions in scientometric data.

1 Assumptions: learning, bounded rationality, evolution

Traditional philosophy of science had considered the institutions of science as given. With stable preferences (textbooks), rationality (the Scientific Method) and equilibrium (truth) the existence of successful science is straightforward. Conceiving of science as a complex system, Kuhn contrasted this traditional view which assumed a given method, clear values and a fixed goal with his own view in which the institutions of science are endogenous, viz. the rules of the game are made as they go along, emerging and declining as a result of the very activity they regulate.⁷ In contrast to a traditional approach assuming stable preferences, rationality and equilibrium Thomas Kuhn developed a “new image of science” (Kuhn 1970, 3) based on learning, bounded rationality and evolution.⁸

Without the strong traditional assumptions which do not explain but simply presuppose scientists’ coordination on the rules of the game, the challenge for Kuhn is to explain how successful coordinated research efforts can emerge from nothing but the local interactions of agents through time. Thomas Kuhn called such successful coordination a *paradigm*. Work within a paradigm is an “attempt to force nature into the conceptual boxes supplied by professional education.” (Kuhn 1970, 5) The function of a paradigm is to allow scientific contributions to be aggregated over scientists and cumulated over time. This “roster of unsolved puzzles” (Kuhn 1970, 184) acts as a standard for the division of labor in science (Kuhn 1977a, 186), a virtual assembly line as it were determining -just like a real assembly line- the expected end result and how the work is parcelled. Like in real factories, coordination on a virtual assembly line allows specialization, a powerful catalyst of scientific progress. A characteristic aspect of Kuhn’s work is that he is interested in the social aspect of science only in as far as it is epistemic. Specialization resulting from scientific coordination is a social factor of science, but conducive to scientific progress. An important difference between the view of social cooperation along social network lines and along virtual assembly lines is that for the latter, just as in real assembly lines, once converted to the same assembly line cooperation among its members does not require social ties but can be mediated through

⁷According to Kuhn there is “a feedback loop through which theory change affects the values which led to that change” (Kuhn 1977a, 336).

⁸“Sketching the needed reconceptualization, I’ve indicated three of its main aspects. First, that what scientists produce, and evaluate is not belief *tout court* but change of belief, a process which I’ve argued has intrinsic elements of circularity, but of a circularity that is not vicious. Second, that what evaluation aims to select is not beliefs that correspond to a so-called real external world, but simply the better or best of the bodies of belief actually present to the evaluators at the time their judgments are reached. [...] And, finally, I’ve suggested that the plausibility of this view depends upon abandoning the view of science as a single, monolithic enterprise, bound by a unique method. Rather, it should be seen as a complex but unsystematic structure of distinct specialties or species, each responsible for a different domain of phenomena” (Kuhn 2000, 119)

exemplars, exemplary contributions which embody the standard for dividing labor. Scientists within the same paradigm can add to each other’s work without knowing each other socially. Note how this interpretation of social ties departs from much of the received literature on social dynamics of science which takes the social network structure of science as a baseline for the diffusion of scientific knowledge.(Sun, Kaur, Milojevic, Flammini, & Menczer 2013)

Interpreting paradigms as virtual assembly lines allows to connect Kuhn’s conceptual framework to citation data and will prove useful further in the paper for operationalizing some of the observable consequences drawn from the model. What does it mean to cite a paper?⁹ In Kuhn’s view, virtual assembly lines emerge from the contributions scientists make to them. This is possible only if scientists have a way of signaling to each other what paradigm their paper is part of. For Kuhn this is the function of citations. He saw citations as a means to anchor a paper in a paradigm. A citation is then not a positive endorsement of the specific content of a paper, but more generally an indication that the cited paper asks the same kind of questions and has similar criteria for what counts as a satisfactory answers. Paradigms might then be observed by detecting overlaps in cited references that cannot be explained by chance alone. The idea of using co-citation networks to observe paradigms and revolutions is itself not new.(Small 2003)¹⁰Very recently Uzzi, Mukherjee, Stringer, and Jones (2013) have developed a quantitative measure based on the analysis of pairwise combinations of references in the bibliography. Any reference pair in a bibliography can be assigned a z-score. Z-scores above zero indicate pairs that appeared more often in the observed data than expected by chance, indicating their conventionality. Conversely a z-score below zero indicates novelty. Using this method, any paper can be ranked on a continuum from explorative to exploitative (or “novel” and “conventional” as they call it) by calculating the median value of the z-scores of the reference pairs in its bibliography.¹¹

Paradigms have been identified as a mode of scientific cooperation that does not require an explicit social tie such as co-authorship or belonging to the same team.¹² Paradigms can then be identified as clusters of papers exhibiting positive z-scores and overlapping reference pairs. For example research at CERN takes place within a very strict division of labor characterized by a well-defined research context. This is what Kuhn calls “normal science”. Normal science is characterized by very clear research questions, methodology and expectations. Papers in such a context are expected to share more common reference pairs than expected by chance. By contrast revolutionary papers that rely less on existing tradition but instead are the basis of a new one are expected to be characterized by non-overlapping reference pairs and negative z-scores. The evolution of paradigms would result in empirically observable shifts or even dissolutions of the overlapping citations pairs in new papers. Even though citation data was all but non-existent in his time, Kuhn is remarkably explicit about this prospect for empirical observation of paradigms and revolutions in scientometric data:

“if I am right that each scientific revolution alters the historical perspective of the community that experiences it, then that change of perspective should affect

⁹For general reviews of the literature on the meaning of citations, see Bornmann and Daniel (2008) and Leydesdorff (1998).

¹⁰For a different approach to observing scientific revolutions in scientometric data, see Marx and Bornmann (2013).

¹¹For another application of Uzzi et al. (2013), Lee, Walsh, and Wang (2015) use its method to test the effects of team size and variety on creativity

¹²For an analysis of novelty and impact in teams, see Lee et al. (2015)

the structure of postrevolutionary textbooks and research publications. One such effect -*a shift in the distribution of technical literature cited in the footnotes to research reports*- ought to be studied as a possible index to the occurrence of revolutions.” (Kuhn 1970, xi, my italics)

2 Agent-Based Model

A paradigm is a state of coordination over what the questions should be and what counts as a solution. This state of coordination emerges from the individual contributions of scientists. The more scientists coordinate on the same standard, the larger the paradigm. The rise and fall of a paradigm can then be conceived as the dynamics of adoption of a standard, analogous to the dynamics of adoption of technological standards. (Arthur 1989, De Langhe & Greiff 2010) Paradigms emerge and dissolve as a result of the collective action of individual scientists. Kuhn failed to spell out how the different aspects of his image of science were connected. He was never able to specify exactly how successful science as we know it emerges from the individual-level interactions between scientists as Kuhn had characterized them.

“Even those who have followed me this far will want to know how a value-based enterprise of the sort I have described can develop as a science does, repeatedly producing powerful new techniques for prediction and control. To that question, unfortunately, I have no answer at all [...] The lacuna is one I feel acutely” (Kuhn 1977a, 332-3)

To understand how paradigms can possibly emerge and dissolve from the interactions of individual agents requires an agent-based approach which was unavailable at the time Kuhn wrote *Structure*. In this section I specify an incentive structure based on just two general, qualitative assumptions and heuristic rules for individual decision making.¹³ I show that these two very general assumptions are sufficient to produce the aggregate patterns Kuhn described. This is done while remaining within the confines of the Kuhnian assumptions described in the previous section.

2.1 Learning: exploration and exploitation

If standards are not universal but learnt, then multiple standards are possible. As a result the incentive structure for individual scientists is governed by the fundamental dilemma between dividing labor over the standard they know or looking for new and potentially better standards. This tradeoff between exploration and exploitation plays a central role in any organizational learning process (March 1991). Thomas Kuhn (1977b) himself called this the “essential tension” between tradition and innovation.¹⁴ According to Kuhn this essential

¹³See also ? for an analysis of this model.

¹⁴“This central role of an elaborate and often esoteric tradition is what I have principally had in mind when speaking of the *essential tension* in scientific research. I do not doubt that the scientist must be, at least potentially, an innovator, that he must possess mental flexibility, and that he must be prepared to recognize troubles where they exist. That much of the popular stereotype is surely correct [...] But what is no part of our stereotype and what appears to need careful integration with it is the other face of this same coin. We are, I think, more likely fully to exploit our potential scientific talent if we recognize the extent to which the basic scientist must also be a firm traditionalist” (Kuhn 1977b, 239, my italics)

tension between exploration and exploitation is key both for the incentive structure within which agents operate as well as the heuristics they use to navigate it:

“we must seek to understand how these two superficially discordant modes of problem solving can be reconciled both within the individual and within the group.” (Kuhn 1977b, 239)

The incentive for exploiting the same standard for division of labor is *specialization*.¹⁵ United on the same standard scientists can:

“pursue selected phenomena in far more detail, designing much special equipment for the task and employing it more stubbornly and systematically” (Kuhn 1970, 18)

“So long as the tools a paradigm supplies continue to prove capable of solving the problems it defines, science moves fastest and penetrates most deeply through confident employment of those tools. The reason is clear. As in manufacture so in science - retooling is an extravagance to be reserved for the occasion that demands it.” (Kuhn 1970, 76)

The incentive for exploration is *innovation*. If no scientist ever explored new possible standards for dividing scientific labor, science would lock in to a (potentially suboptimal) standard and scientific progress would be impossible. The ongoing tension between exploitation and exploration is the result of the fact that exploitation causes exploration and vice versa: “research under a paradigm must be a particularly effective way of inducing paradigm change.” (Kuhn 1970, 52)

The dynamic tension between exploration and exploitation is modeled as follows. Consider a community of $N(1, \dots, n)$ scientists. Each turn, each scientist makes a contribution $C(c_1, \dots, c_N)$ to a paradigm $S(s_1, \dots, s_M)$; note that N is a constant of the system, however, M may vary as the system evolves. To reconstruct the incentive structure characterized by the essential tension between exploration and exploitation I will make two very simple but general qualitative assumptions, namely that the utility of making a contribution to a paradigm increases with adoption (because of specialization) and decreases with production (because of diminishing innovativeness).

Adoption A of a paradigm s at time t is the sum of the number of scientists that contribute to it at time t .

$$A_s(t) = \sum a_{i,s}(t). \quad (1)$$

¹⁵Kuhn illustrates the benefits of specialization for the electricians using the Franklinian paradigm: “Freed from the concern with any and all electrical phenomena, the united group of electricians could pursue selected phenomena in far more detail, designing much special equipment for the task and employing it more stubbornly and systematically than electricians had ever done before. Both fact collection and theory articulation became highly directed activities. The effectiveness and efficiency of electrical research increased accordingly” (Kuhn 1970, 18)

Production P of a paradigm s is the sum of all contributions ever made to that paradigm. Assuming for simplicity that all agents make a single contribution each turn production is the sum of adopters through time t :

$$P_s(t) = \sum_{t'=0}^{t'=t} A_s(t') \quad (2)$$

Note that the relation between adoption and production as defined in the model reproduce the tension between exploration and exploitation. A contribution to a paradigm will increase (in the short term) and decrease (in the long term) the value of contributions made to it. Hence research under a paradigm is both attractive but at the same time effectively inducing paradigm change.

If the parameter α denotes the output elasticity of coordination, the utility of the next contribution to a paradigm within this incentive structure can then be expressed as:

$$U_s(t) = \frac{(A_s(t) + 1)^\alpha}{P_s(t) + 1}. \quad (3)$$

A paradigm is a standard for dividing labor in science. Division of labor enables specialization and hence scientific progress. The more scientists adopt the same standard for dividing labor, the more specialization is possible. The parameter α represents the increase in specialization allowed by an increase in adoption. Not all domains of scientific inquiry allow for the same amount of specialization, because the technology for standardization has not yet been developed (e.g. biology before the invention of the microscope) or because their ontology is too diverse. More ontological homogeneity makes it easier to specialize, just as more generally standardization of tools and components is a condition of possibility for division of labor. In this model this parameter is therefore a property of the entire community. For the purpose of this paper α is exogenous. The value of α affects the incentive to exploit.

2.2 Bounded rationality: persuasion game

Confronted with the incentive structure characterized in the previous section, individual agents face a decision problem between exploiting their current paradigm or exploring a new one. “The successful scientist must simultaneously display the characteristics of the traditionalist and of the iconoclast.” (Kuhn 1977b, 227) This entails three important consequences for individual decision making.

First, the decision between exploitation and exploration is analytically intractable because the choice set, all possible paradigms, is undefined. Moreover agents cannot evaluate standards directly against each other (*incommensurability*) because their specific characteristics are themselves what is used for evaluation: “the choice is not and cannot be determined merely by the evaluative procedures characteristic of normal science, for these depend in part upon a particular paradigm, and that paradigm is at issue.” (Kuhn 1970, 94) These characteristics of the individual decision process Kuhnian agents face have prompted many critics to declare that Kuhnian agents must be irrational¹⁶ and hence science for Kuhn is only a matter

¹⁶See for example Popper (1970), Scheffler (1982) and Shapere (1984)

of “mob psychology” (Lakatos 1970) or “a political and propagandistic affair” (Laudan 1977). Yet this early wave of criticism was probably exaggerated. In many domains even in our daily lives we make successful decisions under uncertainty. We manage to do this not by brute-force calculations but on the basis of *heuristics*. (Gigerenzer 2000) A heuristic only tells agents how to look, not what to find. Thus it guides the decision process without determining it. It is less specified than an algorithm, but it is this lack of specificity which makes it robust against choice for unknown alternatives. As such, heuristics are not inferior to algorithms, but a different solution to a different problem. In response to his critics Kuhn therefore specified that in his view paradigm choice is based on heuristic values rather than on an algorithm, which he described as “*criteria that influence decisions without specifying what those decisions must be.*” (Kuhn 1977a, 330) He named five: accuracy, consistency, scope, simplicity and fruitfulness. Although paradigms cannot be compared against each other using paradigm-specific attributes, the exploration/exploitation incentive structure nevertheless allows to value them because its evaluation is based on non-paradigm-specific attributes, viz. attributes that any paradigm will have regardless of its content: adoption, production and an output elasticity of coordination. Kuhn’s five heuristic values can be projected on the variables. Adoption can account for accuracy because adoption determines the extent of division of labor and hence the extent of specialization. Production can account for consistency and simplicity because production determines the amount of articulation of a paradigm. The output elasticity of coordination can account for scope and fruitfulness. This projection of Kuhn’s values onto the essential tension between exploration and exploitation (adoption and production) captures Kuhn’s suggestion that values conflict. (Kuhn 1977a, 321)

Apart from incommensurability scientists also face practical limitations. They do not have perfect information. Tracking ongoing work in their paradigm is a costly investment. Scientists cannot simply be assumed to track ongoing work in other paradigms as well. In fact, whether or not to start tracking ongoing work in other paradigms is part of the question under consideration when choosing between exploration and exploitation. For this reason I will assume that scientists only know adoption, production and output elasticity of coordination for their own paradigm (this captures Kuhnian incommensurability) in their own (Moore) neighborhood (this captures Kuhn’s suggestion that scientific values are weighed differently in different contexts. (Kuhn 1977a, 321))

Finally, decision-making by scientists acting within the exploration/exploitation incentive structure is forward-looking. Not backward-looking choosing among existing theories, but forward-looking possibly also searching for new theories. They must find “the fittest way to practice future science” (Kuhn 1970, 172) decide based not on information about previous contributions but based on the *expected* value of their own prospective contribution to a paradigm. What counts is therefore not the value of the last contribution to a paradigm in terms of adoption and production, but that of the next one (current adoption and production +1).

Kuhn described shifts in paradigm allegiance as a conversion experience driven by the efforts of individual scientists to persuade each other. (Kuhn 1970, 152) Within the exploration/exploitation incentive structure it makes sense to want to persuade others to join one’s paradigm because (at least in the short term) the benefit of increased specialization outweighs the disadvantage of increased production. The incentive structure characterized in the previous section

and the three considerations in this section allow to make a precise specification of a Kuhnian “persuasion game”. Assume each agent tries to persuade one of its 8 (Moore) neighbors each turn. Despite the incommensurable, perspectival and forward-looking nature of paradigm choice, precise probabilities can be assigned to any possible action in any possible context of such actions by others: 1) persuasion fails; the target sticks to its paradigm i (P_s), 2) persuasion succeeds; the target is converted to paradigm j (P_c) and 3) persuasion mutation; the targeted is converted to a new paradigm k (P_n).

$$P_s = \frac{U_{s_i}}{U_{s_i} + U_{s_j} + 1}, \quad (4)$$

$$P_c = \frac{U_{s_j}}{U_{s_i} + U_{s_j} + 1}, \quad (5)$$

$$P_n = \frac{1}{U_{s_i} + U_{s_j} + 1}. \quad (6)$$

Each agent only knows the value of the next contribution to its own paradigm. For this reason values are not compared directly and agents cannot simply choose that contribution with the highest value. Rather the probability of persuasion depends on the relative value of each alternative. Put intuitively, the persuasion power of an agent depends on how valuable the next contribution to its paradigm will be relative to the next contribution of the target’s paradigm and creating a new paradigm. For example if an agent adopting a paradigm to which the next contribution is worth 66 within the exploration/exploitation incentive structure tries to convert an agent adopting a paradigm which for that agent is worth 33, the probability of conversion is 66%, the probability conversion fails is 33% and the probability that a new paradigm emerges is 1%. Note that there is a non-zero probability that a new paradigm is created. As a result paradigms are endogenous. The number of paradigms in the model is not specified in advance but a function of the interaction of agents as the model runs. The endogeneity of the number of paradigms is a powerful feature of this model. In fact in the exploitation/exploration incentive structure the value of a contribution to a new paradigm is always exactly 1 because a new paradigm has no adoption ($A = 0$) and no production ($P = 0$). So the next contribution will bring both adoption and production to 1, which according to equation 3 always results in a value of $1/1 = 1$ regardless of α . As a consequence the probability that an entirely new paradigm emerges from a persuasion attempt will vary inversely proportional to the value of existing paradigms. Thus the system self-regulates to find a dynamic balance between the exploitation of existing paradigms and the exploration of new paradigms.

2.3 Evolution

In pseudocode, the model runs as follows.¹⁷

SETUP:

Two-dimensional grid of size $N = L^2$

Let each patch be a scientist

Let the color of each patch represent the paradigm to which that scientist currently contributes

Create N scientists

Let all scientists make a contribution to a randomly selected paradigm

RUN:

Ask N randomly assigned agents to be persuaded by one of their neighbors with the probability of success (conversion), fail and mutation given by 4, 5 and 6. Agents only change paradigm if they are persuaded by an adopter of another paradigm to do so.

All scientists make one contribution to the paradigm of their choice; their color changes accordingly.

Figure 1 shows a few typical runs of the model for various values of α . A few observations can already be made. Paradigms as geographically concentrated pockets of coordination emerge, move, decline, split up and disappear again. The structure of the community changes as α increases. The number of paradigms decreases as the incentive for exploitation (α) increases. Although there is no limit to the creation of new paradigms, paradigms are only created as existing ones are exhausted. The model thus self-organizes to find a dynamic balance between exploration and exploitation. Although these patterns appear to live a life of their own, they nevertheless emerge exclusively from the local interactions of boundedly rational individuals based on their locally available information.

3 Observable consequences

In section 1 I have introduced paradigms as virtual assembly lines that emerge together with the contributions scientists make to them. In the previous section I have constructed an agent-based model of how paradigms can emerge and decline as a consequence of rational interactions between scientific agents choosing what assembly line to join. This section deals with the question how these assembly lines and their evolution could in principle be observed in scientometric datasets. Although other scientometric data might be relevant, for this paper I will limit myself to citation data.

Thomas Kuhn distinguished between those areas of science in which there continue to be a wide range of paradigmatic assumptions (“pre-paradigmatic science”) and “mature” areas of science in which some paradigmatic assumptions have become more widely adopted than others. In mature sciences, dominance of paradigms is punctuated by periods of revolutionary science. Philosophers like Toulmin (1970) argue that without a principled way to distinguish between them, Kuhn’s entire model is jeopardized. The model of the dynamics of scientific revolutions introduced in the previous section makes two in principle testable predictions: a polarization between little specialized and highly specialized communities and

¹⁷The model was written using the *Netlogo* software package version 4.1.3

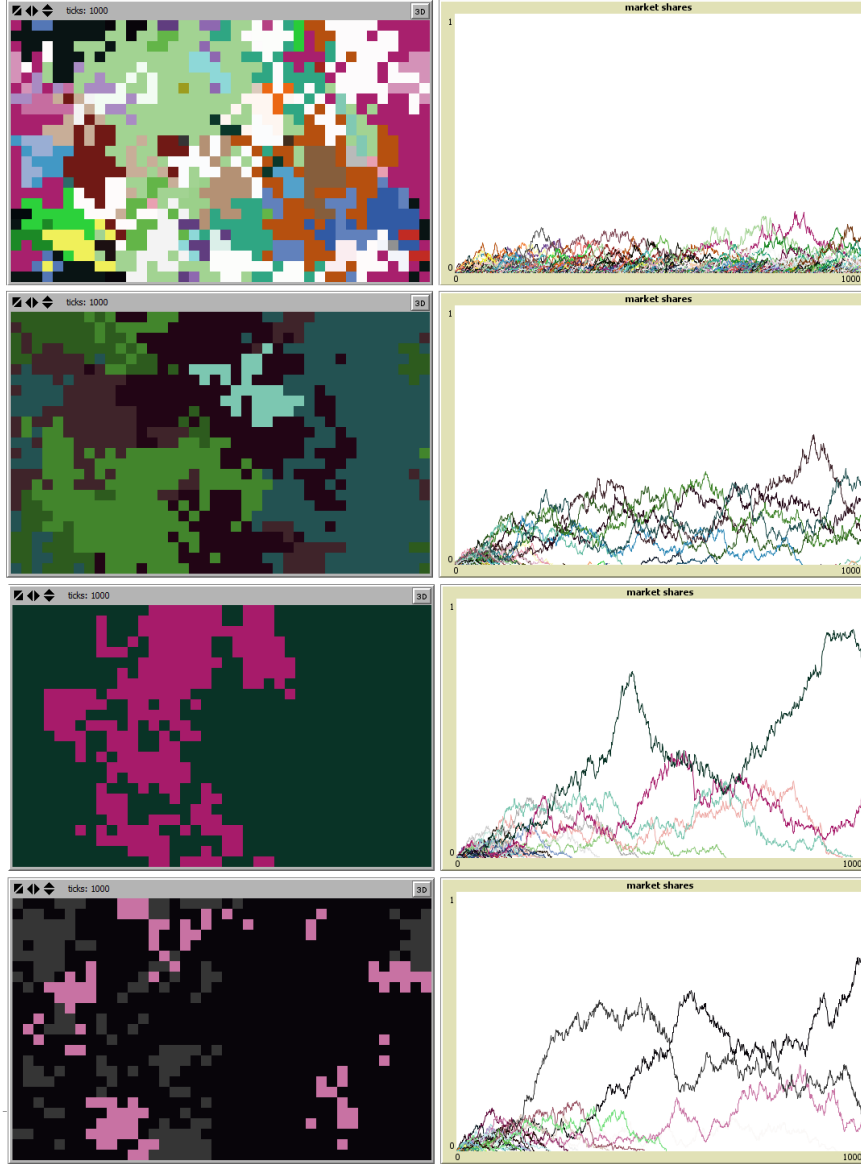


Figure 1: Typical state of the model after 1,000 turns (left) with evolution of market share throughout the run (right) for (from the top down) $\alpha = 2, 5$, $\alpha = 5$, $\alpha = 7, 5$ and $\alpha = 10$.

within these highly specialized communities an alternation between periods of high exploitation punctuated by periods of exploration. By explaining how patterns can emerge from the interactions of individual agents, agent-based models allow us to better understand these patterns and potentially even derive previously unexpected but in principle observable consequences. (Bonabeau 2002, Railsback & Grimm 2011) Here I show that the agent-based model in the previous section makes two empirical predictions by which the distinction between pre-paradigmatic and mature science on the one hand and normal and revolutionary science on the other could be empirically observed in scientometric data.

3.1 Pre-paradigmatic and mature science

The first consequence is that specialization in a community on a shared assembly line occurs quite suddenly as α increases. Variation of all sizes can be observed around $\alpha = 6.5$, perhaps suggesting a critical point. Communities with α below this point are characterized by the continuous presence of multiple competing paradigms, entailing a low level of specialization. Communities in which α is higher are characterized by the alternating monopoly of a single paradigm punctuated by shorter periods of crisis. Fig.2 shows that the average share of scientists making a contribution to the dominant paradigm in their community exhibits a cross-over from very low to very high as α increases. Around the tipping point, communities characterized by a minor difference in α are expected to exhibit large differences in structure. The different runs for various community sizes L^2 illustrate that this tipping point for α is independent of the size of the community. Assuming a uniform random distribution of α across scientific domains this entails a polarized scientific landscape with on the one hand communities characterized by very low levels of specialization and on the other hand communities characterized by very high levels of specialization, with little in between. More precisely, in terms of Uzzi et al. (2013) (see section 1), a polarized distribution of Z-scores is expected across science. Deviations from this prediction might be explained by variations in the ontological homogeneity of the domains studied by scientists (or a preference among scientists to study certain domains rather than others). Deviations can also be explained as a result of the dataset. For example Uzzi et al. (2013) use Web of Science data. However Web of Science makes inclusion in their database conditional on a number of requirements as a result of which only journals from more specialized communities are included.

3.2 Normal and Revolutionary Science

The second consequence that can be derived from the agent-based model introduced in the previous section is a difference in the distribution of the total number of contributions made to dominant paradigms (normal science) and the total number of contributions made to non-dominant paradigms (revolutionary science). Fig.3 shows how often it occurred that paradigms have a total production of a certain size. The x-axis denotes the total number of contributions C to paradigms S , the y-axis how often this number occurred $P(C/S)$ during the run of the model (100,000 ticks). This was done for three different values of α , one for pre-paradigmatic science ($\alpha = 5$), one for mature science ($\alpha = 8$) and one around the tipping point ($\alpha = 6.5$). The top row shows the distribution of total production, the second row shows the distribution of production in non-dominant paradigms and the third row shows the distribution in dominant paradigms.

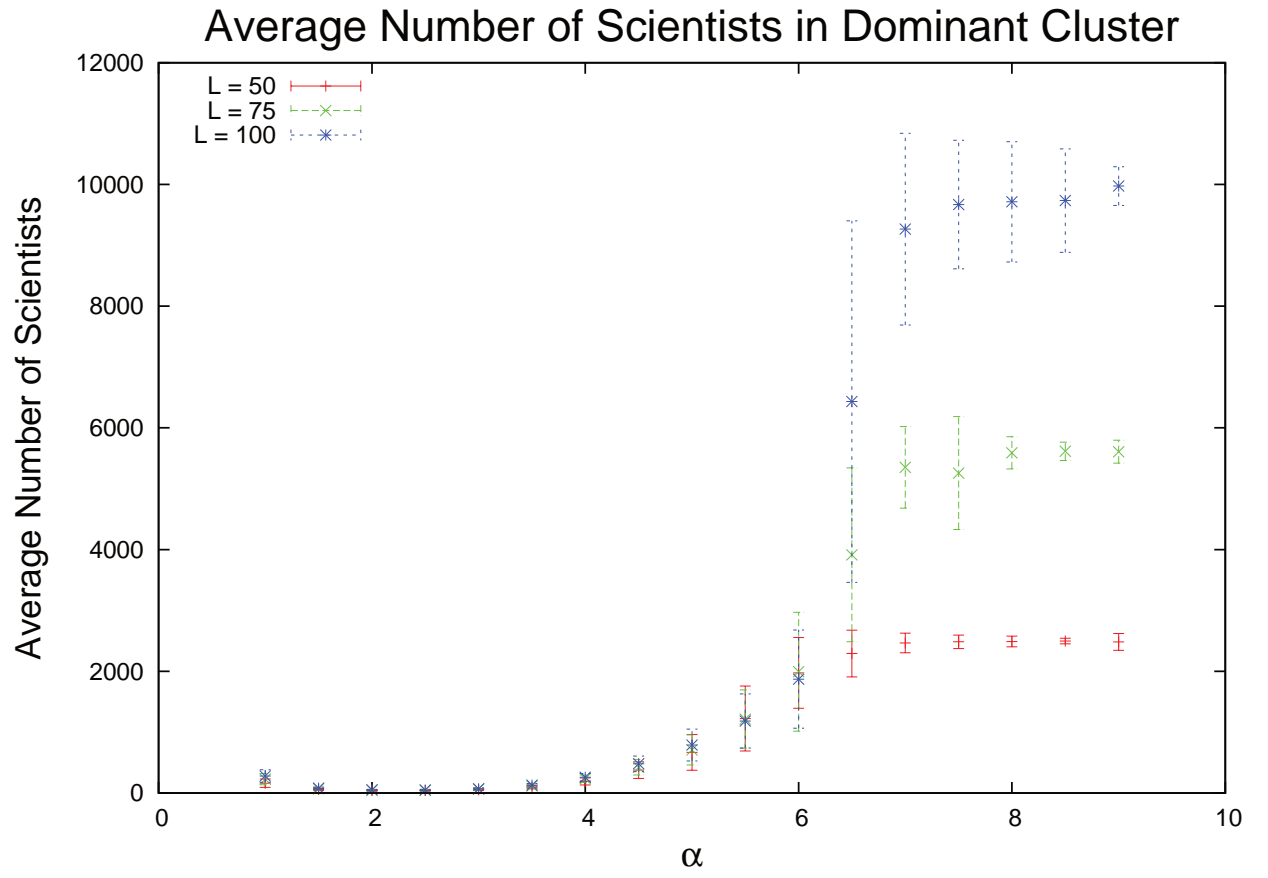


Figure 2: Average number of adopters to the dominant paradigm for various α and different dimensions L (100,000 ticks).

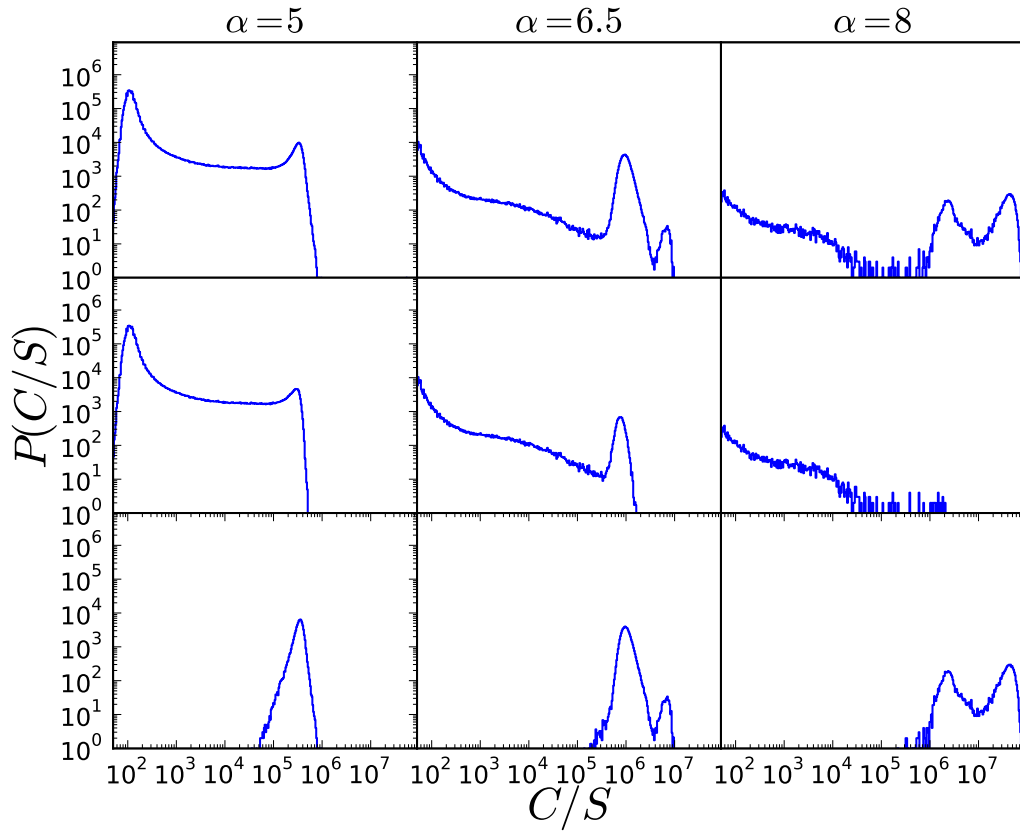


Figure 3: Top row shows distribution of total production for $\alpha = 5, 6.5, 8$; $L = 100$ (100,000 ticks). Middle row shows distribution of production to non-dominant paradigms (revolutionary science), lower row to dominant paradigms (normal science). Note how the difference in distribution between normal and revolutionary science becomes more pronounced as α increases.

The distribution of total production for various values of α (top row) shows a change in the distribution of total production as communities evolve from pre-paradigmatic to mature science. For mature science ($\alpha = 8$) the distribution of total production appears to consist of two separate distributions, suggesting a difference in the distribution of total production between normal science and revolutionary science. This suggestion is reinforced by the fact that separating the distribution of total production to dominant and non-dominant paradigm in the rows underneath is sufficient to disentangle both distributions. Revolutionary science is characterized by a distribution of total production of a wide range of sizes. But these sizes are typically smaller than normal science, which is characterized by a distribution that shows less variation in but higher numbers of production. A possible explanation for this difference in size variation is that the size of non-dominant paradigms is limited by each other's various sizes, while the size of a dominant paradigm is limited only by the (in this model fixed) size of the community. A possible explanation for the higher total production is the observation that periods of revolutionary science are typically more fragmented and shorter than periods of normal science.

Detection of these different statistical distribution would support the distinction between normal and revolutionary science. The model predicts that total production in normal and revolutionary science are characterized by statistically different properties, allowing for a quantitative separation of these phases based on the distribution of their total production. In terms of Uzzi et al. (2013) normal science papers are characterized by high Z-scores while revolutionary science papers have lower Z-scores. Empirical observation will require a good proxy for the total number of contributions to a paradigm. Assuming that the total number of citations of a paper is a proxy for the total production of the paradigm it was contributed to, a comparison of the distribution of the total number of citations to papers with high Z-scores and papers with low Z-scores could be used to test for the occurrence of a double distribution in mature science. A difference in distribution would mean there are indeed two separate phases, one normal and one revolutionary.

References

- Andersen, H., Barker, P., & Chen, X. (2006). *The cognitive structure of scientific revolutions*. Routledge.
- Arthur, B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *Economic Journal*(394), 116-131.
- Arthur, B. (1994). *Increasing returns and path-dependence in the economy*. Ann Arbor: University of Michigan Press.
- Bettencourt, L., Kaiser, D., & Kaur, J. (2009). Scientific discovery and topological transitions in collaboration networks. *Journal of Informetrics*, 3, 210-221.
- Bettencourt, L., & Kaur, J. (2011). Evolution and structure of sustainability science. *PNAS*, 108(49), 19540-19545.
- Bettencourt, M., & Kaiser, D. (2015). Formation of scientific fields as a universal topological transition. *SFI Working Paper*(2015-03-009). (Retrieved from <http://www.santafe.edu/media/workingpapers/15-03-009.pdf>)
- Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(3), 7280-7.

- Bornmann, L., & Daniel, H.-D. (2008). What do citation counts measure? *Journal of Documentation*, 64(1), 45-80.
- Chen, C. (2012). *Turning points*. Berlin: Springer.
- Chen, C., Chen, Y., Horowitz, M., Hou, H., Liu, Z., & Pellegrino, D. (2009). Towards an explanatory and computational theory of scientific discovery. *Journal of Informetrics*, 3(3), 191-209.
- De Langhe, R. (2013). The kuhnian paradigm. *Topoi*, 32(1), 65-73.
- De Langhe, R. (2014). A unified model of the division of cognitive labor. *Philosophy of Science*, 81(3), 444-459.
- De Langhe, R., & Greiff, M. (2010). Standards and the distribution of cognitive labour. *Logic Journal of the IGPL*, 18(2), 278-294.
- Fuller, S. (2001). *Thomas kuhn: A philosophical history for our times*. Chicago: Chicago University Press.
- Garfield, E. (1964). *The use of citation data in writing the history of science*. Philadelphia: Institute of Scientific Information.
- Gigerenzer, G. (2000). *Simple heuristics that make us smart*. Oxford: Oxford University Press.
- Holland, J. (1998). *Emergence from chaos to order*. Oxford: Oxford University Press.
- Hoyningen-Huene, P. (1993). *Reconstructing scientific revolutions*. Chicago: Chicago University Press.
- Kuhn, T. (1970). *The structure of scientific revolutions, 2nd ed.* Chicago: Chicago University Press. ([1962])
- Kuhn, T. (1977a). *The essential tension*. Chicago: Chicago University Press.
- Kuhn, T. (1977b). The essential tension: Tradition and innovation in scientific research. In *The essential tension* (p. 225-239). Chicago: Chicago University Press. ([1959])
- Kuhn, T. (2000). *The road since structure: Philosophical essays 1970-1993*. Chicago: Chicago University Press.
- Lee, Y.-N., Walsh, J., & Wang, J. (2015). Creativity in scientific teams: unpacking novelty and impact. *Research Policy*, 44(3), 684-697.
- Leydesdorff, L. (1998). Theories of citation? *Scientometrics*, 43(1), 5-25.
- March, J. (1991). Exploration and exploitation in organizational learning. *Organization Science*, 2(1), 71-87.
- Marx, W., & Bornmann, L. (2013). The emergence of plate tectonics and the kuhnian model of paradigm shift: A bibliometric case study based on the anna karenina principle. *Scientometrics*, 94(2), 595-614.
- Masterman, M. (1970). The nature of a paradigm. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (p. 59-89). Chicago: Chicago University Press.
- Mazloumian, A., Eom, Y. H., Helbing, D., Lozano, S., & Fortunato, S. (2011). How citation boosts promote scientific paradigm shifts and nobel prizes. *PLoS ONE*, 6(5).
- Miller, J., & Page, S. (2007). *Complex adaptive systems*. Princeton: Princeton University Press.
- Moravcsik, M., & Murugesan, P. (1979). Citation patterns in scientific revolutions. *Scientometrics*, 1(2), 161-169.
- Nickles, T. (2003). *Thomas kuhn*. Cambridge: Cambridge University Press.
- Popper, K. (1970). Normal science and its dangers. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (p. 51-8). Chicago: Chicago University Press.
- Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions:

- Toward an objective measure of scientific impact. *PNAS*, 105(45), 17268-17272.
- Railsback, S., & Grimm, V. (2011). *Agent-based and individual-based modeling*. Princeton: Princeton University Press.
- Scheffler, I. (1982). *Science and subjectivity*. New York: Hackett Publishing Company.
- Shapere, D. (1984). Meaning and scientific change. *Boston Studies in the Philosophy of Science*, 78, 58-101.
- Sharrock, W., & Read, R. (2003). Does thomas kuhn have a 'model of science'? *Social Epistemology*, 17(2-3), 293-6.
- Small, H. (2003). Paradigms, citations, and maps of science: A personal history. *Journal of the American Society for Information Science and Technology*, 54(5), 394-399.
- Sterman, J. (1985). The growth of knowledge: Testing a theory of scientific revolutions with a formal model. technological forecasting and social change. *Technological Forecasting and Social Change*, 28(2), 93-122.
- Sterman, J., & Wittenberg, J. (1999). Path dependence, competition, and succession in the dynamics of scientific revolution. *Organization Science*, 10(3), 322-341.
- Sun, X., Kaur, J., Milojevic, S., Flammini, A., & Menczer, F. (2013). Social dynamics of science. *Scientific Reports*, 3, 1069.
- Toulmin, S. (1970). Does the distinction between normal and revolutionary science hold water? In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (p. 39-47). Chicago: Chicago University Press.
- Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science*, 342, 468-472.
- Wray, B. (2011). *Kuhn's evolutionary social epistemology*. Cambridge: Cambridge University Press.